that $\sigma^2_i = r^2_{ij}\sigma_i^2 + (1 - r^2_{ij})\sigma_i^2$, does not permit us to hold X_j responsible for the share r^2_{ij} of σ^2_i, *unless* X_j is completely independent of all the other causes of X_i, in which case, as Krichewsky shows, $E_{ij} = r^2_{ij}$; in this particular case, *Dines's law* holds, but if r_{ij} is the result of intricate intercorrelation between X_i and a number of mutually correlated causes, then r^2_{ij} merely measures the *degree of covariation* between X_i and X_j; the measure of *causal connection* is E_{ij}. If *all* the X_i are mutually *independent*, and if (2) is *exact*, then $\sum_{i=1}^{n} r^2_{oi} = 1$.

The analysis of the variance of a composite variable by means of the E_{oi}, together with a careful study of the partial correlation coefficients, should be of material assistance in seeking a physical explanation for a series of gross coefficients and in evaluating the relative importance of different causal factors, although there still remains need for caution in drawing final conclusions, particularly (it seems to the reviewer) if $\Sigma E \neq 1$. In this connection, it is helpful to have at hand, for comparison purposes, the relations which hold in various special cases: For example, if three variables exactly satisfy the relation $x_1 = ax_2 + bx_3$, and if $r_{23} = 0$, then if the partial correlation coefficient actually accomplishes what it is supposed to, we should have $r_{12.3} = r_{13.2} = 1.00$; and it is a matter of simple, though somewhat cumbersome, algebra to show that this is the case (*5*); hence $r^2_{12}(1 - r^2_{12}) = r^2_{13}(1 - r^2_{13})$, from which, and the formulæ for the regression coefficients, $r_{12} = a(\sigma_2/\sigma_1)$, $r_{13} = b(\sigma_3/\sigma_1)$; then $E_{12} = (a^2\sigma_2^2)/\sigma_1^2$, $E_{13} = (b^2\sigma_3^2)/\sigma_1^2$; $\sigma_1^2 = E_{12}\sigma_1^2 + E_{13}\sigma_1^2$; $\Sigma\, r^2 = \Sigma\, E = 1$; and $z_1 = r_{12}z_2 + r_{13}z_3$. Again, if M_1 and M_2 are two effects of the cause A_3, $r_{12.3} = 0$, $r_{12} = r_{13}r_{23}$, $r_{13.2} = r_{13}$, $r_{23.1} = r_{23}$; this case has been discussed in some detail by C. F. Marvin in the preceding paper. If A_2, A_3 are the two causes of a result M_1, and are themselves correlated, $r_{12.3} = r_{13.2} = r_{23.1} = 1$. And so on.

As an illustration of how the above principles may be made to aid in the interpretation of correlation coefficients from the viewpoint of cause and effect, Krichewsky applies them to some of W. H. Dines's well-known coefficients; an extended investigation of this character would probably bring out clearly the physical implication of these coefficients and help appreciably in answering the many interesting questions raised by them.

LITERATURE CITED

(*1*) DINES, W. H.
1915. FORECASTING WEATHER BY MEANS OF CORRELATION. Met'l. Mag., 50:30–31.
WALKER, G. T.
1926. ON CORRELATION COEFFICIENTS, THEIR CALCULATION AND USE. Quar. Jour. Roy. Met. Soc., 52:73–84.
(*2*) WHIPPLE, F. J. W.
1924. THE SIGNIFICANCE OF REGRESSION EQUATIONS IN THE ANALYSIS OF UPPER AIR OBSERVATIONS. Quar. Jour. Roy. Met. Soc., 50:237–243.
(*3*) WRIGHT, S.
1921. CORRELATION AND CAUSATION. Jour. Agr. Res., 20:557–585.
KRICHEWSKY, S.
1927. INTERPRETATION OF CORRELATION COEFFICIENTS. Ministry of Public Works, Egypt, Phys. Dept. Pap. 22. Cairo.
(*4*) PEARSON, K.
1911. GRAMMAR OF SCIENCE. Pt. I. 3 ed. London.
(*5*) PHILLIPS, F. M.
1923. APPLICATION OF PARTIAL CORRELATION TO A HEALTH PROBLEM. Public Health Reports, vol. 38, No. 37, pp. 2117–2129. Washington.

A STUDY OF THE POSSIBILITY OF ECONOMIC VALUE IN STATISTICAL INVESTIGATIONS OF RAINFALL PERIODICITIES

By DINSMORE ALTER

[University of Kansas, Lawrence, Kans., December 18, 1926]

In this series of papers embodying a systematic statistical investigation of the world's rainfall, an attempt has been made to refrain from all speculation and to present the evidence so far as possible from the viewpoint of mathematical probabilities of periods versus accidental relationships. For this reason both the causes and the economic value have not been mentioned beyond the briefest discussion several years ago.

It seems wise, however, at the conclusion of the work to make an attempt to learn whether the periodicities found have only a purely scientific interest, or in addition, a possible economic value. Such a value could at the most only pretend to divide seasons in advance into wet, normal, and dry, where wet is defined as including all which average among the wettest third of the data, dry those among the driest third, and normal the remainder. On the basis of accident such predictions should be fulfilled one time out of every three. The work done indicates that *in the long run* such predictions almost certainly can be made with at least a slight increase over this fraction. However, unless the increase is rather large they will have no interest save a purely scientific and statistical one over many years.

To be conclusive, such an investigation must do two things:

(*a*) It must examine the data already available, in order that we may know the percentage of times the periods found will represent the data used in finding them, to this accuracy.

(*b*) It must make *test* predictions that we may follow them through the future and thus weight their value. It is certain that in the long run these can not be fulfilled as accurately as the past representations, for the accidental errors are certain to have modified, more or less, the periods found. In addition, periods of greater or shorter length will have an effect.

It is very important to note that even if we had data which were entirely free from accidental errors and from periods other than those obtained and used in prediction, and even though we knew perfectly the magnitudes and phase relationships of these periods, they would not correctly predict the means for a given stretch of time. In two of the papers of this series, the effect of the datum interval on the magnitude of the amplitude has been investigated and a factor F determined by which multiplication is necessary in order to reduce the amplitude or the intensity found, to what it would have been had much shorter intervals been used. When we have the reverse problem it is necessary that we divide by this factor before predicting. If the predictions are to be made for the same interval used in the original periodogram the factor is eliminated. If not, we must multiply the amplitude obtained from the periodogram by the F corresponding to that ratio of period length to datum interval and divide by that of the ratio to the predicted datum interval. If we do not do this, short periods will exert far too great an influence on our predictions and cause them to fail. In the present preliminary paper, where we are interested not in obtaining the theoretically best predictions possible, but merely in an approximation to them, we shall find that the factor F is only a criterion of length of the periods short enough to be neglected in predictions of means. For such amplitudes as have been found, we shall use about three times the datum interval.

This is six months; therefore we can neglect periods of shorter length than one and one half years. Of course, if a period of length *greater than one year* and of tremendous amplitude with respect to those of the longer periods had existed, it would have been necessary for us to use it, although with extreme caution.

When we examined the periodograms from the halves of the data (MONTHLY WEATHER REVIEW, February, 1927), we found that the correlation in the case of the British Isles was very much higher than in the other cases. Obviously, therefore, whether it may finally prove valuable to predict for any one or all of these sections, our chances of success are far better here than in the other cases. Since the present predictions are not made for actual agricultural use nor even for the purpose of finding whether the rainfall of a given section is predictable but to find whether that of *any* section is so, to an extent which makes it economically valuable, it seems best to confine our study to this one section. If it should prove worth while, these other sections and of course many more must come up for investigation.

In the last paper we showed that the correlation between stretches of the actual rainfall observations separated by 43 years is 3.50 times its probable error and has an expectancy ratio of one in 1,100. Since there were 64 pairs of data to compare we can be quite certain that this is not the result of accident. It may be noted in passing that this relationship is entirely independent of the periodogram or of any theoretical discussion. *It is a matter of actual observation.*

Turning now to the periodograms (this REVIEW, *loc. cit.*) we find that almost without exception the peaks found, even the rather low ones, are through the whole stretch of periods from one and one-sixth to nine years very closely harmonics of a period slightly in excess of this value. This is to be expected after we have found such a high correlation in the data themselves. They of course indicate that if we had compared data with those a trifle more than 43 years later, instead of exactly 43 years, a higher correlation could have been found. We should note that these values were obtained from 75 year's data, a number in no way simply related to 43 years or to 43 and a fraction, and hence in no way open to the objection of being a mere mathematical Fourier series representation.

There are two obvious methods of making our test predictions, each with its advantages and disadvantages. The simpler is to record the data as they occurred 43 years before the predicted date. This is of course very easy to do and to weigh. The other method is to choose those terms which have large amplitude in the periodograms and to combine them, representing both past and future by them. The advantages and disadvantages are:

(*a*) Any small real harmonics of 43 years which are used in the extrapolation, but neglected in the analytical method, improve the predictions by the comparison method.

(*b*) In comparing data themselves, the accidental errors enter into both the old and the new values in such a way that on the average they will produce a discrepancy the $\sqrt{2}$ times as large as in either. If we select only those periods of large amplitude for our work, we shall to a great extent obviate this difficulty.

Which of these two is the better to use for prediction of wet, normal, and dry seasons can not be stated until after a representation has been made by each. First we shall see the results obtained by extrapolation.

In obtaining our correlation coefficient between actual data separated by an interval of 43 years, 64 pairs of values were used. Of these, 38 pairs were either both above or both below normal, three had one value exactly normal and only 23 were of the opposite sign. Excluding the three cases where one value was normal, as being neither an agreement nor a disagreement, we find that the past representation holds 62 per cent of the time and fails 38 per cent. Unlike the case of the analytic representation, it is reasonable to hope that past and future may hold equally well. If so we would be here on the border line of value for economic uses.

It is more useful to consider three groups instead of two. One-third of the data fall in a group 92 to 108 per cent of normal, inclusive. A third fall below 92 per cent and a third above 108 per cent. We find that twenty-eight times out of the sixty-four both values of the pair fall in the same group, a percentage of 44 instead of the 33 expected through accident. It is reasonable to suppose that the pairs containing the oldest data have larger accidental errors than the others. If we consider the latter half of the 64 pairs we find 18 pairs out of 32 in the same group, a percentage of 56 instead of the accidental 33. Quite possibly our correlation would have been higher had we not had larger errors in the 1850–1865 data than in the others. This is quite plainly indicated by the fact that the sum of the squares of the residuals from the normal is largest for these years, although quite evenly balanced between positive and negative values. The larger errors of the earlier data are to be expected not only because of probably less accurate observations but also because we used only three stations instead of five, as in the later part. Possibly we would have improved the correlation had we formed new 6-month means of the second stretch of data, making them fall 43 years and 1 month after the earlier. It does not seem worth while to do it now, for the probable improvement would scarcely affect the decision regarding the test predictions. If predictions for agricultural use were to be made we should, of course, improve them by every such means, even though the gain be slight. The data used end with 1924, therefore the predictions given as Table 1, begin with 1925. They are carried through 1936. If these are fulfilled there will be considerably fewer wet years than we usually have.

In the case where the peaks of the periodograms fall so close to the Fourier harmonics, it seems well to try forcing the curve to the 43-year period shown independently. If we find that a small number of the harmonics represent the dat quite accurately, we shall prefer to use them in making test predictions because of the more balanced manner in which the accidental errors enter. If, on the other hand, many terms are required, we shall use the extrapolation as the method which is probably the more accurate.

Prof. Dayton C. Miller, of Case School of Applied Science, was kind enough to determine the first 30 harmonics of the 43-year period by his harmonic analyser. This new determination was necessary because of the slight discrepancies between periodogram peaks and harmonics, which made some shift in amplitudes and phase relationships.

In order that all the data might be used, values separated by 43 years were averaged. Since the phase relationships must be identical under our hypothesis, such a procedure is certain to give more satisfactory results. They were then plotted as a curve 400 millimeters long.

Twenty of these harmonics showed amplitudes large enough to indicate possible reality. This fact alone indicated clearly that the correlation found was due to many terms. However, a synthesis was made from the eight largest terms. It was no better than that made by extrapolation and, therefore, no predictions are tried from it. Of course all the harmonics could be used for a synthesis which would represent *past* data more accurately than was done.

Seventy-five years' data were used. It may be possible to secure data taken before 1850 sufficiently accurate to use. Eleven years of such data would make it possible to extrapolate from averages of two datum values. If not, the 43-year period will be completed twice by the end of 1935 and any accuracy which may be found for predictions now should be materially increased after that date.

It is quite probable that for any one of three causes the length of the 43-year term may shift as time goes on. These causes have been thoroughly discussed in the previous papers and need only be mentioned here.

In conclusion, a period approximately 43 years, with harmonics, exists in the rainfall data of the British Isles. It may be complex, it may be constant, and it may be variable; time alone can tell which of these is true. Whether predictions made by it at present, or even at some future time, can have economic value is uncertain. However, the chances are sufficient to warrant the attempt, if it be sufficiently emphasized that the predictions are for test purposes only.

About half of the computations for this paper were made under a grant from the Research Committee of the Graduate School of the University of Kansas. I wish also to express my thanks to Professor Miller for the analysis made by him. Without that aid, part of the contemplated work would have remained incomplete on account of lack of time.

TABLE 1.—*Test predictions for British Isles rainfall through 43-year extrapolation*

[A indicates first half of the year, B the second half. Wet, normal, and dry as defined in body of the paper]

	Data A	Data B		Data A	Data B
1925—A	Wet	Wet.	1931—A	Dry	Dry.
B	do	Do.	B	Normal	Normal.
1926—A	Normal	Dry.	1932—A	Dry	Do.
B	do	Normal.	B	do	Do.
1927—A	do	Do.	1933—A	Normal	Dry.
B	Dry	Wet.	B	do	Normal.
1928—A	Normal	Dry.	1934—A	Dry	Do.
B	Dry	Do.	B	Wet	Wet.
1929—A	Wet	Wet.	1935—A	Dry	Dry.
B	Normal	Normal.	B	Normal	Normal.
1930—A	Dry	Dry.	1936—A	Dry	Dry.
B	do	Do.	B	Normal	Do.

ADDENDUM

The additional data received in January, 1927, from the Chief of the Weather Bureau gives two complete 46-year cycles. As a result, predictions should be more accurate, although the probable inaccuracy of observations made nearly a century ago is so great that the gain can not be much. The predictions from the original data are given in the table under Data A, those from the more complete record under Data B. Sixteen predictions are unchanged among 24. Most of the changes are merely shifts across the dividing line between adjacent groups. In only one case, the latter half of 1927, is there any serious shift. For this epoch a prediction, which barely missed being normal, is shifted from dry to just outside of the normal group on the wet side.—*D. A., February 18, 1927.*

THE THUNDERSTORM AT CINCINNATI AND ITS RELATION TO ELECTRICAL POWER SERVICE

By W. C. DEVEREAUX

[U. S. Weather Bureau, Cincinnati, Ohio]

NOTE.—Ninetieth meridian time is used in this article, including the tables and charts. Seventy-fifth meridian time was adopted for Cincinnati by the Weather Bureau on January 1, 1927.

During the 11 years that the Abbe Meteorological Observatory has been maintained in Cincinnati, a most careful and detailed record of all the weather elements has been made. The thunderstorm, like the clouds, must be observed and described by trained observers—no instruments have been devised that will fully meet the requirements or take the place of scientific training. The average recorded number of days with thunderstorms at Cincinnati in the last 10 years shows an increase of 23 per cent as compared with the previous 10 years, due to improvement in the location of the place of observations, and in the methods of observation.

In this study of the individual thunderstorm it is necessary first to define the thunderstorm. Alexander, in his article on the distribution of thunderstorms, has quoted all the instructions to observers on the subject—the last one was issued in January, 1894—and stated that the instructions have been in force ever since. About the only instruction in force at present is that "a storm from which distinct thunder is heard will be considered a thunderstorm," while the instructions to cooperative observers add that "thunderstorms six hours apart may be considered as separate storms."

It is unusual for an individual thunderstorm to last more than two hours, while on the other hand several may occur within one hour. Frequently two or more separate thunderstorms may be visible at the same time. One summer evening distant and diffused lightning was observed in the north for about one hour after dark, then thunder was heard in the north and the separate lightning strokes became visible; other storms started in the west and southwest, until at one time four separate parts of a general storm were visible and thunder could be distinctly heard from each cloud mass—one in the northeast, one in the north, one in the northwest, and one in the west—all storms moving northeastward. Up to this time no rain had fallen at the station, but the next storm, which developed to the southwest, moved directly over the station, and the series of storms that evening were ususually severe. While these storm clouds were separated by a considerable distance, and the path of each was a few miles south of the preceding one, they acted together, and therefore the series should be considered a single thunderstorm. At other times we have observed two or three distinct storms in progress at the same time when there appeared to be very little, if any,